7.1, 7.2 Maximum Likelihood Estimation

continued

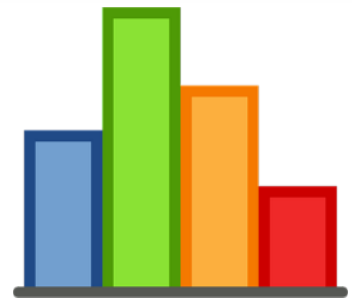
Anna Karlin
Most slides by Alex Tsun

Agenda

- Probability vs Statistics
- Likelihood
- Maximum Likelihood Estimation (MLE)
- MLE Example (Poisson)
- MLE Example (Normal)
- Unbiased estimators

Probability vs Statistics

$Ber(p = 0.5)$ → **Probability** given model, predict data → $P(THHTHH)$



$Ber(p = ???)$ ← **Statistics** given data, predict model ← $THHTHH$

Probability

① Model

$X \sim Poi(5)$

$Y \sim Ber(\frac{1}{3})$

$Z \sim N(0,1)$

② $Pr(X=3) = e^{-5}\frac{5^3}{3!}$

$Pr(Y=1) = \frac{1}{3}$

$Pr(Z \le 3) = \int_{-\infty}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

Likelihood.

① Parametric model

iid. samples from → $Poi(\theta)$, $Ber(\theta)$, $N(\theta_1, \theta_2)$

Use θ to denote parameter(s) of distribution. Unknown to us

Goal: Given iid. samples $(x_1, x_2, \ldots x_n)$

e.g. $Poi(\theta)$

what is most likely value of θ?

Maximum Likelihood Estimation (Poisson)

Let's say $x_1, x_2, \dots, x_n$ are iid samples from $Poi(\theta)$. (might look like $x_1 = 3, x_2 = 5, x_3 = 4$, etc.) What is the MLE of θ?

$$L(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{n} p_X(x_i\,;\,\theta) = \prod_{i=1}^{n} e^{-\theta} \frac{\theta^{x_i}}{x_i!}$$

$$\ln L(\boldsymbol{x} \mid \theta) = \sum_{i=1}^{n} [-\theta + x_i \ln \theta - \ln(x_i!)]$$

$$\frac{\partial}{\partial \theta} \ln L(\boldsymbol{x} \mid \theta) = \sum_{i=1}^{n} \left[-1 + \frac{x_i}{\theta}\right]$$

$$\sum_{i=1}^{n} \left[-1 + \frac{x_i}{\hat{\theta}}\right] = 0 \to -n + \frac{1}{\hat{\theta}} \sum_{i=1}^{n} x_i = 0 \to \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$\frac{\partial^2}{\partial \theta^2} \ln L(\boldsymbol{x} \mid \theta) = \sum_{i=1}^{n} \left[-\frac{x_i}{\theta^2}\right] < 0 \to \text{concave down everywhere}$$

General recipe

Given indep samples $x_1, \ldots, x_n$ from parametric model → Ber(θ) → Poisson(θ)

Find value of θ that maximizes $L(x_1 \ldots x_n | \theta)$

discrete $L(x_1, \ldots, x_n | \theta) \triangleq \prod_{i=1}^{n} p_X(x_i ; \theta)$

cont. $L(x_1, \ldots, x_n | \theta) \triangleq \prod_{i=1}^{n} f_X(x_i ; \theta)$

find $\hat{\theta}$ that maximizes $LL(x_1 \ldots x_n | \theta)$ log likelihood

discrete $= \sum_{i=1}^{n} \ln [p_X(x_i ; \theta)]$

cont $= \sum_{i=1}^{n} \ln [f_X(x_i ; \theta)]$

Find $\hat{\theta}$ to max $LL(\vec{x} \mid \theta)$

$\vec{x}$: $x_1, \dots, x_n$

distn has 1 parameter

compute $\frac{d\, LL(\theta)}{d\theta}$

set $\frac{d\, LL(\theta)}{d\theta} = 0$

solve for $\hat{\theta}$

[verify soln is max (2nd deriv < 0)]

don't need to do this

multiple params $\theta_1, \dots, \theta_k$

find $\hat{\theta}_1, \dots, \hat{\theta}_k$ that are soln to this system

$$\begin{cases} \frac{\partial LL}{\partial \theta_1} = 0 \\ \frac{\partial LL}{\partial \theta_2} = 0 \\ \vdots \\ \frac{\partial LL}{\partial \theta_k} = 0 \end{cases}$$

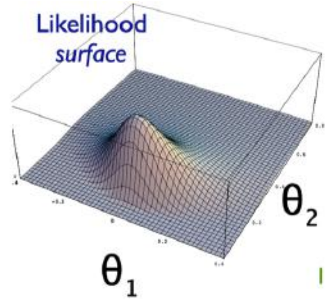
[check max

check Hessian
-ve definite]

Random Picture

Maximum Likelihood Estimation (Normal)

$x_i \sim N(\mu, \sigma^2),\ \mu, \sigma^2$ both unknown



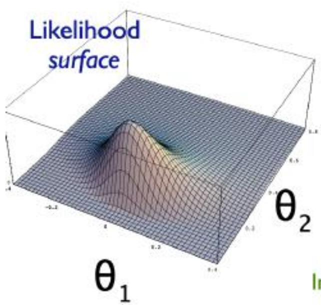
Maximum Likelihood Estimation (Normal)

$x_i \sim N(\mu, \sigma^2)$, μ, σ^2 both unknown

$$\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2}$$

$$\frac{\partial}{\partial \theta_1} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) = \sum_{i=1}^{n} \frac{(x_i - \theta_1)}{\theta_2} = 0$$

$$\boxed{\widehat{\theta}_1 = \left(\sum_{i=1}^{n} x_i\right) / n = \overline{x}}$$



Sample mean is MLE of population mean, again

In general, a problem like this results in 2 equations in 2 unknowns.
Easy in this case, since θ_2 drops out of the $\partial/\partial\theta_1 = 0$ equation

$x_i \sim N(\mu, \sigma^2),\ \mu, \sigma^2$ both unknown

$$\begin{aligned}
\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) &= \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2} \\
\frac{\partial}{\partial\theta_2} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) &= \sum_{i=1}^{n} -\frac{1}{2}\frac{2\pi}{2\pi\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2} = 0 \\
\widehat{\theta}_2 &= \left(\sum_{i=1}^{n} (x_i - \widehat{\theta}_1)^2\right)/n = \bar{s}^2
\end{aligned}$$

Sample variance is MLE of *population* variance